Viser Technology Pte. Ltd.
Mechanical Engineering Science
DOI: 10.33142/mes.v4i2.9086

Research Article Open Access

A CNN-LSTM-PSO tool wear prediction method based on multi-channel feature fusion

Shuo WANG[1], Zhenliang YU[1*], Yongqi GUO[1], Xu LIU[2]

1. School of Mechanical and Power Engineering, Yingkou Institute of Technology, Yingkou, China
2. Yingkou Dingsheng Heavy Industry Machinery Co. LTD, Yingkou, China

***Corresponding Author:** yuzhenliang, email address:yuzhenliang_neu@163.com*

Abstract:
In order to achieve predictive maintenance of CNC machining tools and to be able to change tools intelligently before tool wear is at a critical threshold, a CNN-LSTM tool wear prediction model based on particle swarm algorithm (PSO) optimization with multi-channel feature fusion is proposed. Firstly, the raw signals of seven channels of the machining process are collected using sensor technology and processed for noise reduction; secondly, the time-domain, frequency-domain and time-frequency-domain features of each channel signal are extracted, and a sample data set of spatio-temporal correlation of traffic flow is constructed by dimensionality reduction processing and information fusion of the above features; finally, the data set is input to the CNN-LSTM-PSO model for training and testing. The results show that the CNN-LSTM-PSO model can effectively predict tool wear with an average absolute error MAE value of 0.5848, a root mean square error RMSE value of 0.7281, and a coefficient of determination R2 value of 0.9964; and compared with the BP model, CNN model, LSTM model and CNN-LSTM model, its tool wear prediction accuracy improved by 7.56%, 2.60%, 2.98%, and 1.63%, respectively.
Keywords: feature fusion; CNN-LSTM; tool wear; life prediction

1 Introduction

The severity of tool wear during CNC machining plays a decisive role in the machining accuracy of products, and serious tool wear can reduce product quality, lead to increased scrap rate, and even lead to machine accidents. Therefore, in recent years, tool wear prediction has become a fundamental and prerequisite work in the field of tool life management and intelligent tool change. Early on, experts and scholars have made some progress by exploring the tool wear mechanism and combining Taylor's empirical formula for tool life prediction, the Andis Ābele et al. confirmed the validity of Taylor's empirical formula for predicting tool life and determined the coefficients of Taylor's formula, and finally obtained the formula for predicting the length of the cutting trajectory at the critical wear stage of the tool based on the cutting speed[1] . However, the Taylor's empirical formula only yields a fixed value of tool life, which does not correspond to the actual application of the tool, because the machining parameters are variable and the manufacturing environment is complex, which leads to the impossibility of the remaining tool life in the form of a fixed value.

Based on the above problems, researchers have started to use mechanical learning techniques to predict tool life. Commonly used mechanical learning prediction models are: random forest [2], BP neural network [3], support vector machine (SVM) [4], etc. Wei Weihua [5] et al. optimized BP neural network by genetic algorithm, so that the model's optimization and learning ability can be improved, which can effectively identify tool wear. Sarat Babu Mulpur [6] et al. used OGM-SVM model for real-time prediction of rear tool face wear based on extracted multi-sensor heterogeneous data features and also achieved good prediction results, but the prediction efficiency and accuracy were not high.

In the automated production process, a high accuracy life prediction model can be very effective in predicting the future tool wear level, which is important to study the tool wear at a critical threshold to enable intelligent tool change. Therefore, a large number of experts and scholars have applied deep learning theory in tool life prediction, such as recurrent neural networks (RNN) [7], long and short-term memory networks (LSTM) [8] and convolutional neural networks (CNN) [9], whose prediction effect is significantly higher than mechanical learning techniques. Recently, work on tool life prediction based on long and short term memory networks (LSTM)

Received on October 11, 2022; Accepted on December 3, 2022

has been carried out gradually. Ma Kaile [10]et al. analyzed the singularity of the original vibration signal to eliminate the effect of milling path and constructed a stacked LSTM model for tool wear prediction, and compared with models such as WOA-SVR, it was found that the method improved the accuracy of tool wear prediction. Although the LSTM network can perfectly process the timing information of tool wear, it is difficult to extract the deep features hidden in the samples, which leads to the incomplete extraction of tool wear prediction features and there is still room for improvement.

Convolutional neural networks (CNNs) have strong feature extraction capability and low computational complexity compared with long and short-term memory networks (LSTMs), and can tap deep features hidden in samples. Lim Meng Lip [11] et al. cropped the surface profile images of machined parts and input them into CNN networks for tool wear prediction, and the results showed that the CNN model can meet the tool wear prediction requirements with an accuracy of 98.9 % accuracy. Although these methods have been successful in predicting tool wear, it is still challenging to fully reveal the effective features present in the monitored signals due to the defects in the network structure [12].

As we all know, when the tool wear reaches the sharp wear stage, the system alerts for intelligent tool change, which can improve product machining accuracy and reduce tool management costs. The rule of tool wear is faster in the early stage, slower in the middle stage, and the fastest and most drastic in the late stage. It can be seen that using only one model for tool life prediction will lead to a single extracted feature, which is prone to overfitting. Therefore, combining convolutional neural network (CNN) and long and short-term memory network (LSTM) has become an inevitable trend, using CNN model to extract potential deep features in space and capturing time series information in time by LSTM model, so that the temporal and spatial features of the data can be fully utilized to make up for the shortcomings of the above single prediction model.

In order to further improve the prediction effect of the model, the hyperparameters in the prediction model must be optimized. The more common hyperparameter optimization methods include random optimization [13], gradient-based optimization [14], genetic algorithm optimization [15], particle swarm algorithm optimization [16], etc. The particle swarm algorithm (PSO) can perform global optimization with fewer parameters, and its powerful search performance and individual optimization capability can speed up the convergence of the model, so it has been widely used and studied by scholars in recent years [17].

Therefore, this paper proposes a CNN-LSTM tool wear prediction model with multi-channel feature fusion based on machine vision, feature extraction, deep learning and hyperparameter optimization, constructs a spatio-temporal correlation feature matrix of traffic flow so that the temporal and spatial features of the monitored signal can be fully utilized, and optimizes the hyperparameters in the prediction model using particle swarm algorithm (PSO), so as to improve the tool wear prediction accuracy. The research of this method will propose a new theory and method for tool wear remaining life prediction, and lay a theoretical foundation and scientific basis for improving the development of China's machine tool manufacturing industry and intelligent tool changing field.

2 Construction of CNN-LSTM-PSO prediction model

In order to improve the accuracy and accuracy of the prediction model of tool remaining life, a multi-channel feature fusion CNN-LSTM tool wear prediction model based on particle optimization was proposed in this paper. The output tool wear values were monitored by the vibration signals of three channels, the cutting force signals of three channels and the acoustic emission signals of one channel. Thus, predictive maintenance of NC machining tools can be realized, and tools can be changed intelligently before tool wear is in the critical threshold. The improvements are as follows:

(1) The characteristics of vibration signals, cutting force signals and acoustic emission signals were extracted by batch normalization and dimensionality reduction processing, which improved the generalization ability of the model, avoided overfitting phenomenon and improved the convergence speed of the model.

(2) CNN model reduces network complexity with its unique structure of local connection and weight sharing, and the spatial continuity of sample features is maintained after convolution and pooling operations.

(3) Long term memory network (LSTM) is a further optimization of the traditional RNN network, which can process longer time series data while avoiding the phenomenon of gradient vanishing or gradient explosion.

(4) Using the powerful search and global optimization ability of PSO algorithm, the two parameters of the initial learning rate parameter and the number of hidden layer units in LSTM network were iteratively optimized, which reduced the subjective influence of manual selection parameters, and thus improved the prediction accuracy of tool wear model. The CNN-LSTM-PSO tool wear prediction model is shown in Figure 1.

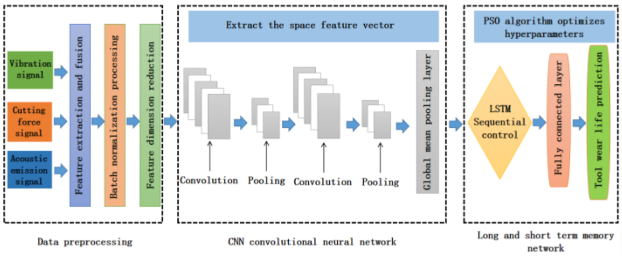


Figure 1 CNN-LSTM-PSO tool wear prediction model

2.1 Convolutional Neural Network (CNN)

Convolutional neural network (CNN) [18] is a kind of neural network, which is a typical representative of deep learning and has obvious advantages for processing spatial data. The most important difference between CNN convolutional neural network and other traditional neural networks is the convolution operation and pooling operation, which can realize local connection and weight sharing. Therefore, the pre-processing part of this paper uses the CNN model to extract the spatial features of the 315×47 sample feature matrix, and its output is a one-dimensional spatial sequence matrix, which lays the foundation for the prediction of tool wear using the LSTM model. The principle is as follows:

(1) The sample feature matrix after batch normalization and dimensionality reduction is input to the CNN convolutional neural network for convolutional operation. The sample information is indirectly characterized by the local features of the sample through the weight value of each layer derived from the convolutional operation, and the higher the layer is, the more detailed the local features are extracted, and also the spatial continuity of the sample is maintained, and its convolutional operation is shown in equation (1):

$$X_i^k = \sum_{j=1}^{n} W_i^{kj} \otimes X_{i-1}^j + b_i^k \tag{1}$$

Where X_i^k denotes the feature matrix of the kth neuron at the output of the ith layer, and W_i^{kj} denotes the weight value of the kth neuron in the ith layer, and $\otimes$ denotes the convolution operator, and X_{i-1}^j denotes the feature matrix of the jth neuron at the output of layer i-1, and b_i^k is the bias coefficient of the kth neuron in layer i.

(2) In order to improve the prediction accuracy of the tool wear life model, the CNN network uses ReLU function for nonlinear activation, which has good non-saturation characteristics to avoid the gradient disappearance phenomenon. The activation function is shown in equation (2):

$$V_i^k = \mathrm{Relu}(X_i^k) = \begin{cases} 0, x_i^k < 0 \\ x_i^k, x_i^k > 0 \end{cases} \tag{2}$$

where x_i^k is theX_i^k each eigenvalue in the feature matrix.

(3) Each tool wear feature data is input to the pooling layer after convolution operation, and the pooling type is selected as maximum pooling, which can retain the original features and reduce the parameters of network training, and improve the robustness of the extracted features. The maximum pooling is shown in equation (3):

$$C_i^k(s,t) = \max_{\substack{1+(s-1)Q \le d \le sQ \\ 1+(t-1)P \le h \le tP}} \{V_i^k(d,h)\} \tag{3}$$

where $V_i^k(d,h)$ is the eigenvalue of column h of row d of the ith feature matrix input to the pooling layer, and $C_i^k(s,t)$ is the eigenvalue of the sth row t column of the ith feature matrix obtained after pooling, and P and Q are the length and width of the pooled region, respectively.

(4) The n feature matrices of dimension S ×T, which are derived from each row of the 315 ×47 sample feature matrix after two convolution and pooling operations, are input to the global average pooling layer. The dimensionality of the pooling kernel of the global average pooling layer is kept consistent with the dimensionality of the feature matrix, and the n feature matrices are dimensionality reduced to reduce the covariance of the sample features and avoid the influence of redundant features, thus reducing the training time of the LSTM long and short term memory network, so the whole CNN model finally outputs a feature vector $X_t = \{x_1, x_2, \dots, x_i, \dots, x_j,\}$ where xi is calculated as shown in equation (4):

$$x_i = \frac{1}{ST}\sum_{s=1}^{S}\sum_{t=1}^{T} C_i^k(s,t) \tag{4}$$

2.2 Long and short-term memory neural network (LSTM)

CNN convolutional neural networks are capable of mining local spatial features related to tool wear, but it is difficult to extract longer time series data. Recurrent neural networks (RNN) can perform temporal processing of tool wear data, but it is difficult to process for longer time series data, and gradient disappearance or gradient explosion occurs during operation. It is usually used to solve this phenomenon using long and short term memory networks (LSTM) or hierarchical RNNs [19]. Long Short-Term Memory Network (LSTM) is a further improvement of the traditional RNN network by introducing memory cells on the input, output, and forgetting past information to construct new cell statesC_t Realize the data transmission, and control the path of data transmission by logic operation through input gate, output gate, and forget gate, so as to complete the processing of longer time series data, and its LSTM network gate cell structure is shown in Figure 2. The new cell state Ct and the output state Ht of the LSTM core are constructed with the following equations:

$$C_t = f_t \otimes C_{t-1} + i_t \otimes \tanh(H_{t-1}) \tag{5}$$

$$H_t = o_t \otimes \tanh(C_t) \tag{6}$$

where f_t is the forgetting gate, which serves to make the cell forget or remember the state of the previous cell C_{t-1} The input gate i_t is the input gate, which controls the input signal and thus updates the memory cell; the current cell state is obtained by reconstructing the cell through the forgetting gate and the input gate C_t The output gateo_t The output gates are used to control the state of the cell C_t The output gates are used to control the state of the cell so that it is transferred to the next cell.

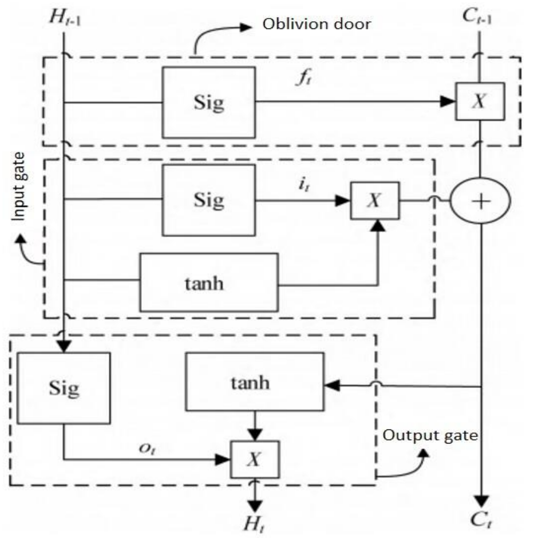


Figure 2 LSTM network gate cell structure

However, the LSTM model also has shortcomings, when dealing with data samples with a large number of features, overfitting is prone to occur, which requires the use of some optimization algorithms to find the optimal number of implied layers and initial learning rate and other parameters to increase the model nonlinear fitting performance and prediction accuracy [20].

2.3 Particle Swarm Optimization (PSO) algorithm

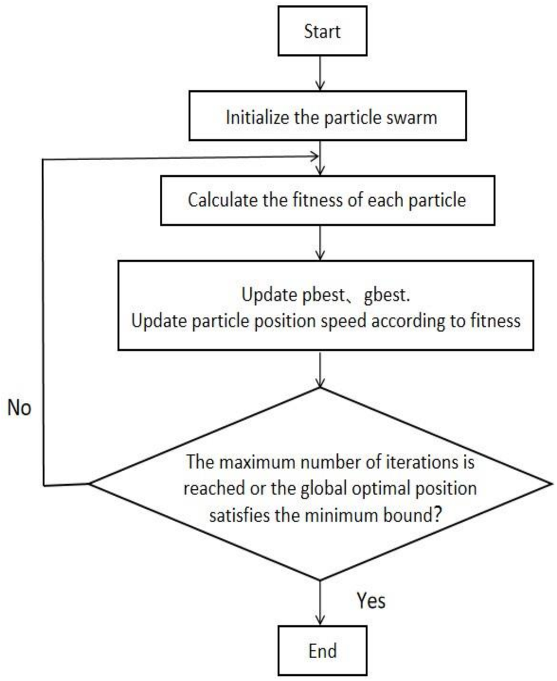


Figure 3 Particle swarm optimization algorithm optimization process

The particle swarm algorithm (PSO) is an intelligent algorithm developed by observing the social behavior of birds. The PSO algorithm is similar to the flock feeding process, and is widely used in the global optimization process of hyperparameters due to its simple principle and easy operation, which refers to the individuals in the population as a particle, and each particle is a possible solution of the optimized parameter in the global search space. Each particle is a possible solution of the optimized parameter in the global search space, and its characteristic index mainly includes three aspects: position, speed and fitness value. Firstly, the fitness value of each particle is calculated by the fitness function to memorize the optimal position and speed of all particles. In each iteration, the particle reaches a new position by adjusting the velocity component of any dimension and calculating it, and so on, until the particle finds the optimal position or reaches the number of iterations, so as to complete the optimization process of the particle in the multidimensional search space, the particle swarm optimization algorithm is shown in Figure 3. In this paper, we use the PSO algorithm to optimize the hyperparameters in the CNN-LSTM model and derive the optimal solution to avoid the overfitting phenomenon during model training.

2.4 CNN-LSTM-PSO hybrid model

In the regression prediction of tool wear, the convolutional layer in the CNN model is first used to obtain the weight parameters, and the pooling layer is used for dimensionality reduction to mine the local features related to tool wear, and its output is a one-dimensional spatial feature vector. The output feature vector is then trained as an LSTM model, which enables the two models to complement each other in time and space, thus improving the accuracy of prediction. Figure 4 shows the CNN-LSTM model prediction process based on particle swarm optimization for multi-channel feature fusion proposed in this paper. The essence is to use the CNN convolutional neural network model as a spatial feature extractor and the LSTM model as a trainer for regression prediction, based on which the superparameters such as initial learning rate and number of hidden layer units in the LSTM model are optimized by the PSO algorithm, so that the model nonlinear fitting performance is improved and the tool wear prediction effect is optimized, and the specific steps are as follows:

Step 1: The original signals of the 7 channels are processed for noise reduction and feature extraction and fusion in the time domain, frequency domain and time-frequency domain, respectively.

Step 2: Using Pearson's correlation coefficient formula to downscale the above feature data to construct the training and test sets of the model.

Step 3: Build a convolutional neural network, train it using the training set and test set from step 2, output a spatial feature vector, and form a new training set.

Step 4: The hyperparameters such as initial learning rate and number of hidden layer units in the LSTM model are used as optimization-seeking processing objects by the PSO algorithm, and the particle swarm optimization model is initialized.

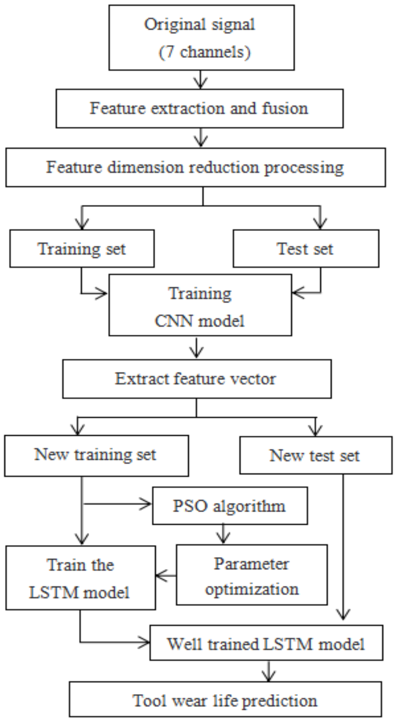


Figure 4 CNN-LSTM-PSO model prediction flow

Step 5: As shown in Figure 6, firstly, the particle's fitness value is calculated, secondly, pbest, gbest are updated according to the fitness, and finally, the position and velocity of the updated particle are recorded.

Step 6: When the maximum number of iterations is reached or the most suitable position is found, the whole loop is terminated and the optimal hyperparameters are derived. If the termination condition is not reached, then return to step 5 for the next iteration.

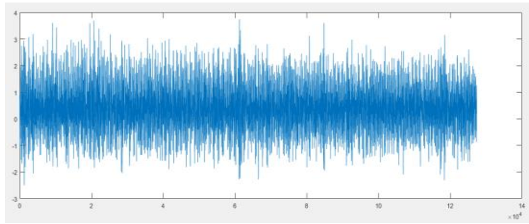
Step 7: Train the LSTM model with the training set formed in step 3 and the hyperparameters obtained in step 6, thus completing the regression prediction of tool wear.

3 Construction of tool wear sample dataset

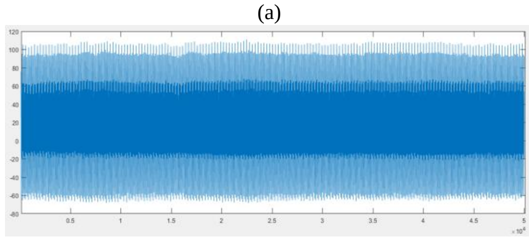
3.1 Multi-channel feature extraction and fusion

The experimental data were obtained from the open data of the 2010 High Speed CNC Machine Tool Health Prediction Contest of the Prediction and Health Management Society (PHM), New York, USA[21]. The dataset is the result of real-time tool wear monitoring experiments on six ball-ended milling tools. In this paper, the experimental dataset of group C1 is selected, where the experimental data of the first 200 tool walks of group C1 is used as the sample training set and the experimental data of the last 115 tool walks are used as the test set. The original signals in each dataset include X-axis, Y-axis and Z-axis cutting force signals, X-axis, Y-axis and Z-axis vibration signals and acoustic emission signals, among which cutting force signals and vibration signals contain 3 channels and acoustic emission signals are 1 channel signals, totaling 7 channels.

In this experiment, the tool is walked once every Δt time, and each time the tool is walked, the original signal of 7 channels can be collected, and the number of collected points of the original signal of each single walk is about 200000 or more, which shows that the number of signal data is huge and there is a lot of noise, and these noises are often caused by the instability of the system at the moment the tool is cut in and out. This requires noise reduction for all types of raw signals collected above to avoid adverse effects during model training. Therefore, the sampling points with data labels from 50001 to 100000 in the raw signal are collected respectively as the research object. The results of the comparison between the original signal and the noise reduction signal are shown in Figure 5, which shows that the signal fluctuation after noise reduction is uniform and noiseless. In this experiment, the number of tool walks in each channel is 315, and there are 7 channels in total, so the original signal can form a 315×7 tool wear signal matrix after noise reduction.



(a)



(b)

Figure 5 Comparison results between the original signal and the noise reduction signal.(a) Raw signal data.(b) Signal data after noise reduction

The 315×7 signal matrix after noise reduction is extracted in the time domain, frequency domain and time-frequency domain, and the time-domain information mainly includes mean, standard deviation, root mean square, etc., totaling 13 time-domain features; the frequency domain information mainly includes frequency domain amplitude mean, center of gravity frequency, mean square frequency, etc., totaling 5 frequency domain features; the wavelet packet decomposition is performed on the original signal, resulting in 8 frequency bands, and the energy of each frequency band is used as time-frequency domain information, totaling 8 time-frequency domain features. The energy of each frequency band is used as time-frequency domain information, and the total is 8 time-frequency domain features, so 26 features can be extracted from each channel signal. The features of all channels are fused to obtain 182 features, and the matrix is reorganized to obtain a 315×182 feature matrix.

3.2 Feature dimensionality reduction processing

The ball-head milling cutter used in this experiment has three teeth in the CNC machining process. In order to improve the accuracy and precision of tool wear prediction, the rear face wear of each tooth needs to be measured and its average value is taken to characterize the actual wear of the tool. In this experiment, there are 315 tool walks, and the average value of the measured wear after each tool walk is composed of a sample target matrix with a matrix dimension of 315×1. Each value in the sample target matrix is the output data of the CNN-LSTM-PSO wear prediction model. In this experiment, the LEICA MZ12 microscope was used to measure the tool rear face wear, and its C1 group tool wear variation curve is shown in Fig. 6, and its variation pattern is consistent with the temporal information mentioned in the previous section.

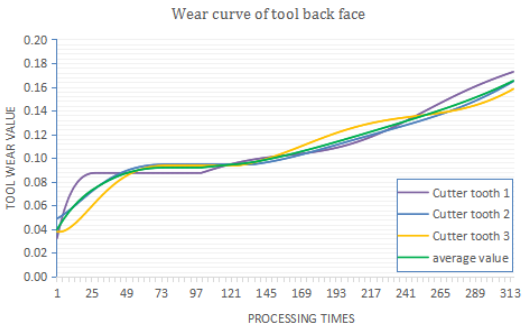


Figure 6 Test tool wear variation curve

According to the above, the extracted features yielded a feature matrix of 315 × 182 by multi-channel feature fusion, but not all the features can characterize the wear of the back tool face. In order to find the correlation between the feature matrix and the target matrix more clearly, the above multi-channel fused feature matrix and the tool wear value are normalized, and the normalized processing formula is

$$X_n = \frac{X - X_{min}}{X_{max} - X_{min}} \tag{7}$$

The correlation between the normalized sample data and the tool wear curve is shown in Figuer 7. It can be seen from the figure that there are many features that do not correlate with the tool wear values or have weak correlation that will interfere with the tool wear prediction model and should be given to be removed. And Pearson correlation coefficient is the most widely used correlation coefficient analysis method, which can be used to measure the correlation between the extracted feature values and tool wear [22]. Its calculation formula is:

$$P_{xy} = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{\sqrt{n\sum x_i^2 - (\sum x_i)^2}\sqrt{n\sum y_i^2 - (\sum y_i)^2}} \tag{8}$$

where P_{xy} denotes the Pearson correlation coefficient of the signal feature x and the tool wear value y. The Pearson correlation coefficient formula is used to calculate the 315 × 182 feature matrix and filter out $|P_{xy}| \geq 0.9$ the strongly correlated features as the input of the prediction model. In total, 47 strongly correlated features are extracted through the calculation, taking the X-axis cutting force signal as an example, 7 strongly correlated features are obtained after dimensionality reduction, and the sample data of the signal after dimensionality reduction are shown in Figure 8, which shows that the noise signals with poor correlation are deleted, and the stripped out data with poor correlation are shown in Figure 9; in this paper, the 47 strongly correlated features are fused and reorganized, and the dimensionality of the sample feature matrix is 315×47. 47, and this sample feature matrix is the input data of CNN-LSTM-PSO wear prediction model.

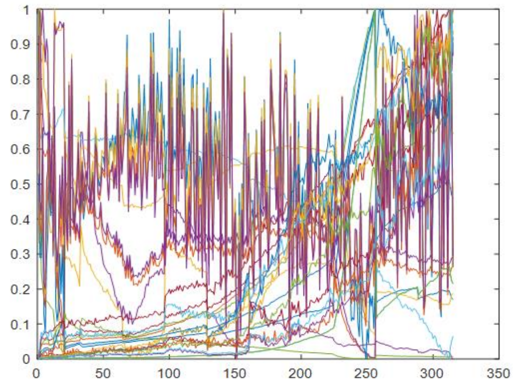


Figure 7 Normalised sample dataset

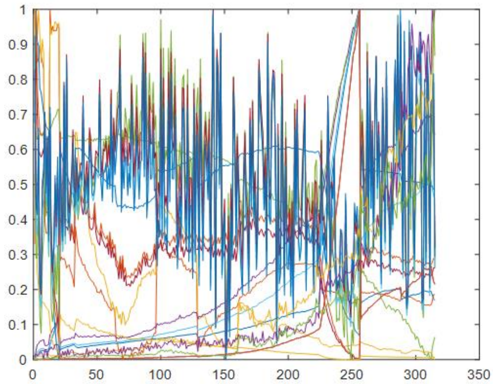


Figure 8 Sample data after dimensionality reduction

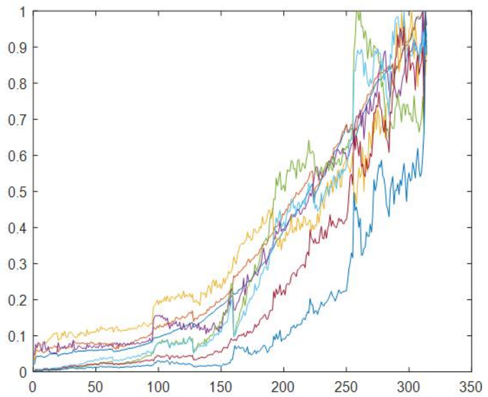


Figure 9 Deleted poor correlation data

4 Experimental verification and analysis of tool wear

4.1 Tool wear experimental conditions

The experimental conditions for tool wear are shown in Figure 10, whose cutting vibration signals were collected using a Kistler 8636C piezoelectric accelerometer, cutting force signals were collected using a Kistler 8152 three-way platform dynamometer, and acoustic emission signals were collected using a Kistler 9265B acoustic transmitter, whose relevant CNC machining cutting parameters are shown in Table1.

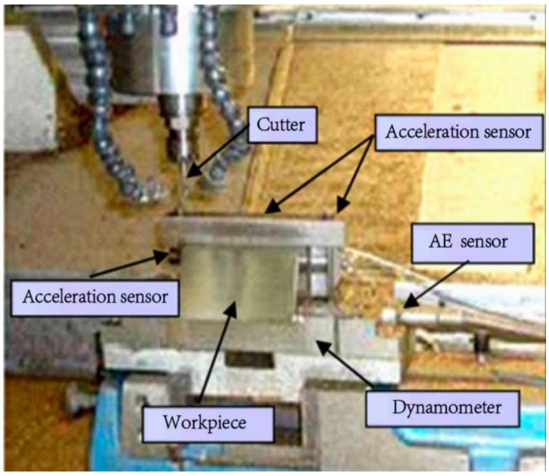


Figure 10 Experimental conditions for tool wear

Table 1 CNC machining cutting parameters

Main shaft Rotational Speed	Feeding Speed	Back draft	Side Eating Knife quantity	Feed amount	Cold cutting conditions
10400	1555	0.2	0.125	0.001	Dry cutting

In this paper, the raw signals related to tool wear are collected in real time according to the above experimental conditions, and each channel raw signal is processed by noise reduction, extraction, fusion and dimensionality reduction to obtain a 315×47 sample feature matrix, and a sample dataset with spatio-temporal correlation of traffic flow is jointly constructed with 315×1 sample target matrix with dimensionality of 315×48. CNN-LSTM-PSO The model first inputs the sample dataset into the multilayer CNN model to extract the spatial sequence features of the traffic flow data and outputs the spatial feature vector. Then the spatial feature vector is input to the multilayer LSTM model to extract the time series features of the data, thus combining the temporal features and spatial features. Finally, the PSO algorithm is used to optimize the hyperparameters in the CNN-LSTM model, so as to complete the prediction of tool wear.

4.2 Setting of prediction model parameters

In order to avoid the influence of external factors, the number of particle swarm individuals in the PSO algorithm is set to 15 and the maximum number of iterations is set to 60. The values of the initial learning rate parameter of the optimized CNN-LSTM model are set between 0.001 and 0.01, and the values of the number of hidden layer units are set between 1 and 100. The structural parameters of the tool wear prediction model after hyperparametric optimization based on the PSO algorithm are shown in Table 2.

Table 2 Structural parameters of CNN-LSTM-PSO model

Structural section	Network structure Name	Parameter settings
1	Convolutional layer 1 Batch standardisation layer 1 Pooling layer 1	Activation function: RELU Convolution kernel: 3*3 Maximum pooling
2	Convolutional layer 2 Batch standardisation layer 2 Pooling layer 2	Activation function: RELU Convolution kernel: 3*3 Maximum pooling
3	LSTM layer 1	Learning rate: 0.004 Number of hidden layer units: 50 Activation function: Sigmoid
4	LSTM layer 2	Learning rate: 0.004 Number of hidden layer units: 32 Activation function: Sigmoid
5	Dropout layer	25% discard
6	Output layer	Activation function: Softmax

In order to quantify the prediction performance of the tool life model, three objective evaluation indexes are selected, namely the mean absolute error MAE, the root mean square error RMSE and the coefficient of determination R2. Among them, the mean absolute error MAE can obtain an evaluation value, but the comparison between different models is required to reflect the model's superiority; the mean square error RMSE and the coefficient of determination R2 can directly characterize the model's superiority. The smaller the mean square error RMSE and the closer the coefficient of determination R2 is to 1, the higher the accuracy and precision of the prediction model. The three evaluation indicators are calculated as follows:

$$MAE = \frac{\sum_{t=1}^{m}|y_t - \hat{y}_t|}{m} \quad (9)$$

$$RMSE = \sqrt{\frac{\sum_{t=1}^{m}(y_t - \hat{y}_t)^2}{m}} \quad (10)$$

$$R^2 = 1 - \frac{\sum_{t=1}^{m}(y_t - \hat{y}_t)^2}{\sum_{t=1}^{m}(y_t - \bar{y})^2} \quad (11)$$

where, m is the number of samples output from the fully connected layer, the number of samples in this paper is 315, and $\hat{y}_t$ is the predicted value of tool wear, and y_t is the actual value of tool wear.

4.3 Tool life prediction results

In this paper, a CNN-LSTM model with multi-channel feature fusion using particle swarm optimization is used for tool wear regression prediction, and its test set prediction results are shown in Figure 11. The mean absolute error MAE value of the model was calculated to be 0.5848, the root mean square error RMSE value was 0.7281, and the coefficient of determination R2 value was 0.9964. The results show that the use of CNN-LSTM-PSO based model can effectively perform regression prediction of tool wear and achieve better results.

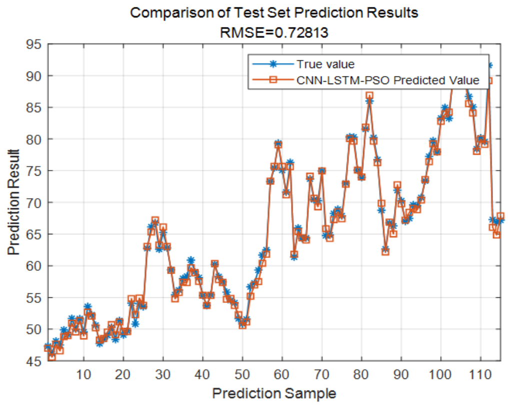


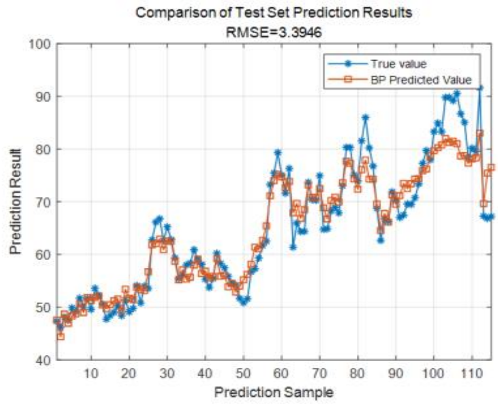
Figure 11 CNN-LSTM-PSO test set prediction results

Table 3 shows the effect of the PSO algorithm on the tool wear regression prediction model, where the hyperparameters such as the initial learning rate and the number of hidden layer units of the CNN-LSTM model rely on manual random selection, and it can be seen that the CNN-LSTM model optimized using the PSO algorithm has the best tool wear prediction. Compared with the CNN-LSTM model, its mean absolute error MAE and root mean square error RMSE are reduced and the coefficient of determination R2 is improved, and its performance index exceeds 0.99, while the performance index of the CNN-LSTM model with manually selected parameters is maintained at a maximum of about 0.98. This is mainly because the PSO algorithm obtained more accurate hyperparameter pairings after hyperparameter optimization of the CNN-LSTM model, which found the most critical attributes affecting the accuracy of tool wear prediction and avoided the blindness of setting parameters, thus improving the prediction results.

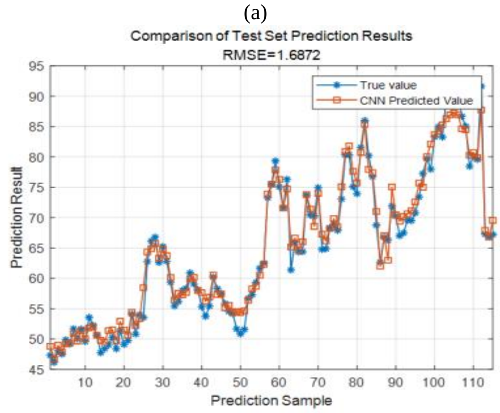
Table 3 Effect of PSO algorithm on prediction model

Algorithm	Initial learning rate	Number of hidden layer units		Test set prediction results		
		LSTM 1	LSTM 2	MAE	RMSE	R2
CNN-LSTM	0.01	100	50	2.9757	3.5829	0.9128
	0.01	60	20	2.2307	3.0005	0.9388
	0.001	100	50	2.0172	2.1781	0.9675
	0.001	60	20	0.9718	1.1914	0.9802
CNN-SVM-PSO	0.004	50	32	0.5848	0.7281	0.9964

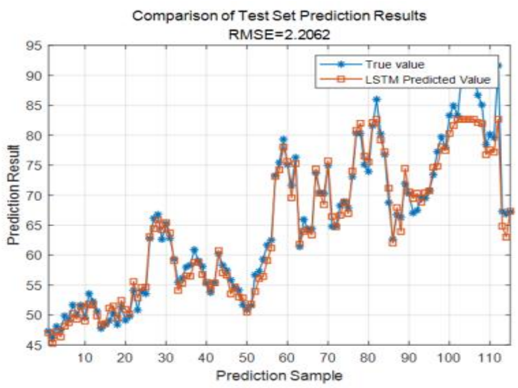
To further validate the prediction performance of CNN-LSTM-PSO based tool wear, a comparative analysis was performed with other traditional prediction models in the past, such as BP neural network, CNN model, LSTM model and CNN-LSTM model. Figure 12 shows the comparison results of the four traditional tool wear prediction models, and it can be seen from Figure 11 that the root mean square error RMSE values of the CNN-LSTM-PSO model proposed in this paper are reduced by 78.59%, 56.85%, 66.99%, and 38.89% compared to the BP model, CNN model, LSTM model, and CNN-LSTM model, respectively. This shows that the prediction performance of the CNN-LSTM tool wear prediction model optimized based on the PSO algorithm proposed in this paper is superior due to the single algorithm of other traditional prediction models, incomplete feature extraction, and over-reliance on signal processing techniques and expert experience.



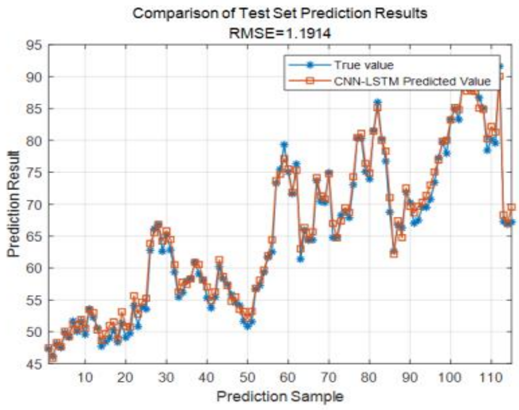
(a)



(b)



(c)



(d)

Figure 12 Prediction results of the four traditional models

(a) BP model. (b) CNN model. (c) LSTM model.(d) CNN-LSTM model.

Table 4 Comparison of prediction performance results of five models

Algorithm	Test set prediction results		
	MAE	RMSE	R2
BP Neural Network	2.5413	3.3946	0.9211
CNN Algorithms	1.3242	1.6872	0.9705
LSTM Algorithms	1.5425	2.2062	0.9667
CNN-LSTM Algorithm	0.9718	1.1914	0.9802
CNN-SVM-PSO Algorithm	0.5848	0.7281	0.9964

Table 4 shows the comparison results of the prediction performance of the five models. It is found that the CNN-LSTM-PSO model using multi-channel feature fusion for tool wear prediction has the smallest value of mean absolute error MAE, which is reduced by 76.98%, 55.84%, 62.09% and 39.82% compared to the BP, CNN, LSTM and CNN-LSTM models, respectively ; the value of the coefficient of determination R2 is closest to 1, which is 7.56%, 2.60%, 2.98%, and 1.63% higher compared to the BP, CNN, LSTM, and CNN-LSTM models, respectively. These two results again prove that the prediction of tool wear values using the CNN-LSTM-PSO model proposed in this paper is more accurate and can achieve more effective monitoring of remaining tool life and intelligent tool change.

5 Conclusion

In this paper, the open dataset of the tool health prediction competition is selected as the original data, and the original data is preprocessed using feature extraction and multi-channel fusion techniques, and then a CNN-LSTM model based on particle swarm optimization with multi-channel feature fusion is proposed to predict the tool wear values during milling machining, and compared with other single mechanical models and the traditional CNN-LSTM model analysis, and the results show that:

(1) In this paper, the CNN model is used to extract local features from the feature matrix after multi-channel fusion and dimensionality reduction to obtain important information of tool wear data and avoid the interference of tool wear data by other factors.

(2) The parameter search optimization of the tool wear prediction model by the particle swarm PSO algorithm reduces the subjective influence of manual parameter selection and avoids the blindness of setting parameters, thus improving the prediction accuracy.

(3) Tool wear regression prediction using the CNN-LSTM-PSO model has a mean absolute error MAE value of 0.5848, a root mean square error RMSE value of 0.7281, and a coefficient of determination R2 value of 0.9964. This indicates that the model can effectively predict the remaining life of the tool with good results.

(4) Compared with the BP model, CNN model, LSTM model and CNN-LSTM model, the mean absolute error MAE and root mean square error RMSE values of the CNN-LSTM-PSO model proposed in this paper have been reduced, and the value of the coefficient of determination R2 has been improved to be closest to 1. This indicates that the constructed tool life prediction model has less error, better accuracy and better.

In the future, the CNN-LSTM-PSO tool wear prediction model can be widely used in the fields of intelligent tool change and tool life management for CNC machining in various factories. By predicting the tool wear value in real time, it can realize intelligent tool change in advance when the tool wear is at the critical threshold, thus improving the machining accuracy of products.

Author Contributions: For Conceptualization, methodology, analysis, and writing original draft preparation, Wang Shuo; writing review and full-text editing, Yu Zhenliang; writing—original draft preparation, Guo Yongqi; writing—original draft preparation, Liu Xu.

Conflicts of interest: The authors declare no conflict of interest.All authors have read and agreed to the published version of the manuscript.

Acknowledgments: The research work financed with the means of Basic Scientific Research Youth Program of Education Department of Liaoning Province, No.LJKQZ2021185; Yingkou Enterprise and Doctor Innovation Program (QB-2021-05).

References

[1] Andis Ābele, Henn Tuherm. Predictions of Cutting Tool Wear of Straight Milled Aspen Wood with Taylor's Equation [J]. Current Journal of Applied Science and Technology, 2016(5):7.

[2] Cynthia Deb, M. Ramesh Nachiappan, M. Elangovan, V. Sugumaran. Fault Diagnosis of a Single Point Cutting Tool using Statistical Features by Random Forest Classifier [J]. Indian Journal of Science and Technology, 2016, 9(33):45-46.

[3] Xu Yanwei, Gui Lin, Xie Tancheng. Intelligent Recognition

Method of Turning Tool Wear State Based on Information Fusion Technology and BP Neural Network [J]. Shock and Vibration, 2021(2):55-56.

[4] Alajmi Mahdi, Almeshal Abdullah. Estimation and Optimization of Tool Wear in Conventional Turning of 709M40 Alloy Steel Using Support Vector Machine (SVM) with Bayesian Optimization [J]. Materials,2021, 14(14):23.

[5] Wei Weihua, Cong Rui, Li Yuantong. Prediction of tool wear based on GA-BP neural network [J]. Proceedings of the Institution of Mechanical Engineers,2022, 236(12):8-9.

[6] Sarat Babu Mulpur, Babu Rao Thella. Multi-sensor heterogeneous data-based online tool health monitoring in milling of IN718 superalloy using OGM (1, N) model and SVM [J]. Measurement, 2022(199):723-724.

[7] Lee Hojin, Jeong Hyeyun, Koo Gyogwon. Attention RNN Based Severity Estimation Method for Interturn Short-Circuit Fault in PMSMs [J]. IEEE Transactions on Industrial Electronics, 2020(5):7.

[8] Han Sung-Ryeol, Kim Yun-Su. A fault identification method using LSTM for a closed-loop distribution system protective relay [J]. International Journal of Electrical Power and Energy Systems, 2023(148):5-8.

[9] Zhou Yuankai, Wang Zhiyong, Zuo Xue. Identification of wear mechanisms of main bearings of marine diesel engine using recurrence plot based on CNN model [J]. Wear, 2023(6): 520-521.

[10] Ma Kaile, Wang Guofeng, Yang Kai. tool wear monitoring for cavity milling based on vibration singularity analysis and stacked LSTM [J]. The International Journal of Advanced Manufacturing Technology, 2022(120):5-6.

[11] Lim Meng Lip, Derani Mohd Naqib, Ratnam Mani Maran, Yusoff Ahmad Razlan. tool wear prediction in turning using workpiece surface profile images and deep learning neural networks [J]. The International Journal of Advanced Manufacturing Technology, 2022(120):11-12.

[12] Jiahang L, Xu Z. Convolutional neural network based on attention mechanism and BiLSTM for bearing remaining life prediction [J]. Appl Intell. 2021(52):1076 -1091.

[13] Stefan Droste, Thomas Jansen, Ingo Wegener. Upper and Lower Bounds for Randomized Search Heuristics in Black-Box Optimization. Electron. Colloquium Comput [J]. Complex, 2003(77):48-48.

[14] Weifeng Lu, Bingyu Cai, Rui Gu. Improved Particle Swarm Optimization Based on Gradient Descent Method [J]. CSAE, 2020(2): 121-126.

[15] Salih Omran, Duffy Kevin Jan. Optimization Convolutional Neural Network for Automatic Skin Lesion Diagnosis Using a Genetic Algorithm [J]. Applied Sciences, 2023,13(5):7.

[16] Zhang Xin, Jiang Yueqiu, Zhong Wei. Prediction Research on Irregularly Cavitied Components Volume Based on Gray Correlation and PSO-SVM [J]. Applied Sciences,2023,13(3):87-89.

[17] Wang Ji, Zhou Jian, Mo Wen-An. Tool life prediction based on multi-source feature PSO-SVR neural network [J]. Journal of Physics: Conference Series, 2022,2366(1):754-756.

[18] Gajera Himanshu K., Nayak Deepak Ranjan, Zaveri Mukesh A.. A comprehensive analysis of dermoscopy images for melanoma detection via deep CNN features [J]. Biomedical Signal Processing and Control, 2023,79(2):46-50.

[19] Ning Zhang, Enping Chen, Yukang Wu, et al. A novel hybrid model integrating residual structure and bi- directional long short- term memory network for tool wear monitoring [J].The International Journal of Advanced Manufacturing Technology , 2022(120):6707-6722

[20] Li Xianwang, Qin Xuejing, Wu Jinxin, et al. tool wear prediction based on convolutional bidirectional LSTM model with improved particle swarm optimization [J]. The International Journal of Advanced Manufacturing Technology, 2022,123(11-12):89-92.

[21] Huimin Chen. A Multiple Model Prediction Algorithm for CNC Machine Wear PHM [J]. International Journal of Prognostics and Health Management, 2011,2(2):78-89.

[22] Li Yifan, Xiang Yongyong, Pan Baisong, et al. A hybrid remaining useful life prediction method for cutting tool considering the wear state [J]. The International Journal of Advanced Manufacturing Technology, 2022(121):5-6.